Use of Neural Networks in a Search for Single Top Quark Production at DØ

Lev Dudko, for the DØ Collaboration

Institute of Nuclear Physics, Moscow State University

Abstract. We present a search for electroweak production of single top quarks in the DØ detector at the Tevatron collider. After initial selections, the signal forms less than one percent of the background, and requires a powerful analysis tool to separate it from background. For this purpose, we employ the neural network package MLPfit, and train it on Monte Carlo (MC) models of two processes for signal, and on data and MC models of five processes for background. Based on an analysis of singularities in the Feynman diagrams for single-top production, we choose an optimal set of kinematic variables as inputs to the networks. We use separate networks for each signal-background pair. For the dominant backgrounds, we use sequential nets.

INTRODUCTION

After the discovery of the top quark in $t\bar{t}$ pair production via the strong interaction (1), the next natural step is to test its properties and the parameters of top physics. One of the most interesting way to do that is to examine the electroweak production of single top quarks (2). This production mode can provide directly the magnitude of the CKM element V_{tb}, the width of the top quark, and certain spin effects resulting from the (V–A) vertex. Beyond the standard model, we can test the Wtb couplings for anomalous behavior, and search for FCNC couplings or W' production. The rate for single top quark production is sufficiently large to pose a significant background to the search for the Higgs boson and other new physics processes. Single top quarks can be produced at the Tevatron in two main processes, via s-channel and t-channel mechanisms (Fig. 1), with cross sections of:

$$\sigma_{\mathrm{NLO}}(p\bar{p}\to t\bar{b}(\bar{t}b)X) = 0.73 \pm 0.04\,\mathrm{pb} \qquad (3)$$
$$\sigma_{\mathrm{NLO}}(p\bar{p}\to tq\bar{b}(\bar{t}qb)X) = 1.70 \pm 0.19\,\mathrm{pb} \qquad (4)$$

We perform our search in the leptonic decay channels of the W boson from top quark decay, with $W\to e\nu_e,\ \mu\nu_\mu$, and $W\to\tau\nu_\tau\to e\nu_e\bar{\nu}_\tau\nu_\tau,\ \mu\nu_\mu\bar{\nu}_\tau\nu_\tau$. This signal is mimicked by several background processes (2):

$$\sigma_{\mathrm{LO}}(p\bar{p}\to Wb\bar{b}, Wc\bar{c}, Wcs, Ws\bar{s}) = 35\ \mathrm{pb}$$
$$\sigma_{\mathrm{LO}}(p\bar{p}\to Wjj,\ j = u,d,g) = 966\ \mathrm{pb}$$
$$\sigma_{\mathrm{NLO}}(p\bar{p}\to WW, WZ) = 13\ \mathrm{pb}$$
$$\sigma_{\mathrm{NNLO-NNLL}}(p\bar{p}\to t\bar{t}) = 6.5\ \mathrm{pb}$$

QCD multijet events with misidentified leptons

Background rates are huge relative to signal, and different backgrounds have essentially different kinematics. These are the reasons to use the neural network (NN) technique in the analysis. After several attempts to improve the signal-to-background ratio, we have developed the NN strategy described below, and train each signal process versus each kind of background.

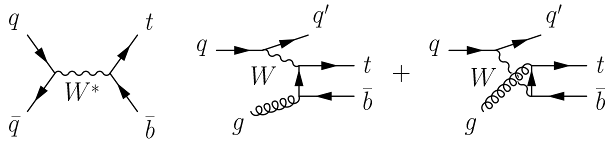

FIGURE 1. Representative Feynman diagrams for single top quark production.

APPLYING THE NN TECHNIQUE

Training the Neural Networks.

The first step in applying the NN technique for separating signal from background involves making a proper choice of input variables. A useful observation that helps to choose an optimal set of input variables is that the rate for a scattering process is greatest in the regions of phase space near singularities of the corresponding matrix element. If such singularities occur in different places for signal and background, then the dependence of the corresponding variables should differ markedly between signal and background. In general, there are two types of singularities: s-channel, $M^2_{f1,f2} = (p_{f1}+p_{f2})^2$, and t-channel $\hat{t}_{i,f} = (p_f - p_i)^2$. Here p_f is the four momentum of the final particle and p_i of the initial particle (parton). The position of the singularity in each of these variables is determined by the denominator of the corresponding Feynman propagator. We use this idea to choose an optimal set of kinematic variables (5).

Before implementing the final NN analysis, we choose to restrict all event samples by eliminating regions of phase space where background is large and expected signal small. In these regions, the kinematic characteristics of background and signal differ greatly, and such events can therefore be easily separated, and the background removed at essentially any stage of the analysis. (In our specific analysis, the excluded regions also correspond to regions where the background tends to be poorly modeled.) The background that remains after such preprocessing, being closer in character to signal, is therefore more difficult to distinguish from signal. However, with sufficient acceptance, the small loss of signal from such preprocessing is not debilitating, and the NN can then be trained using the most sensitive variables for separating signal and backround. An example of the NN outputs for such preprocessed and not preprocessed events is given in Fig. 2, and shows the impact of the preprocessing. For the training of our networks we use

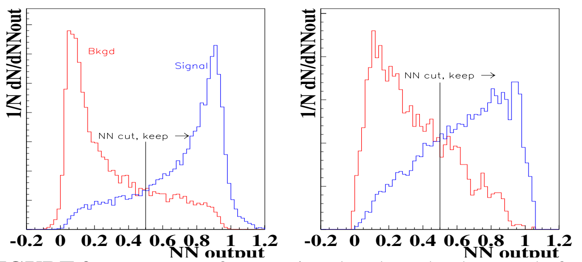

FIGURE 2. NN outputs for one signal and one background after looser (left plot) and tighter (right plot) preprocessing selection.

the MLPfit package (6), with our own code interface and a Hybrid Linear-BFGS training method. We obtain best results using the structure of the NN shown in Fig. 3. The numbers in the boxes correspond to the numbers of nodes on the input, hidden and output layers. We use a structure of three subnetworks for each of the Wjj and Wbb backgrounds. In this structure, we combine two different kinds of input information to improve the extraction of signal from these backgrounds. In addition to kinematic information, we use muon-tagging information for b-jets. In the first subnetwork (Wjj NN1), we combine the muon tagging information and kinematic information. In the second subnetwork (Wjj NN2), we use only kinematic information. We then combine the outputs of the first and second subnetworks into a third one. As shown in Fig. 4, this mix of b-tag information and kinematics leaves more signal and less background than is possible in the simpler structure: e.g., a cut on

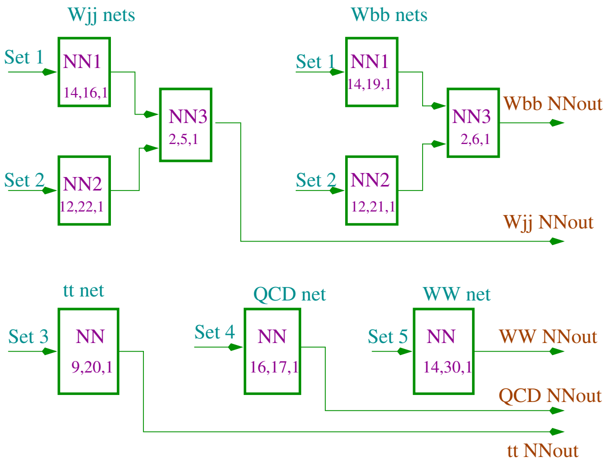

FIGURE 3. Neural network architecture.

only the first or on the second subnetwork (horizontal or vertical line on the plot) in comparison to that on the NN3 output. For the $t\bar{t}$, WW, and QCD backgrounds,

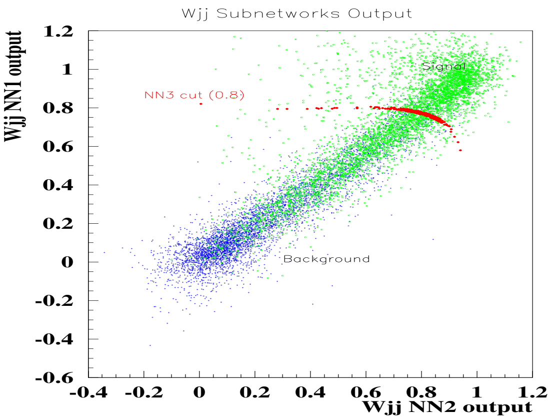

FIGURE 4. Output of the first Wjj subnetwork, versus that of the second. The cut on the third subnetwork is shown.

we train one network for each to separate each from signal. The outputs of each of these networks for signal and background are shown in Fig. 5, with the distributions normalized to unity. To perform the analysis, we pass all data, background and signal samples, through all networks in parallel, and cut on the output of each of the networks.

Check of the NN

To check the stability of the NN analysis, we perform two tests of the networks. In the first step, during training, we divide the input sample (signal and background) into two parts. We train the NN on one of these parts, and at every cycle of training we check for improvement of the training using the test sample, as shown in Fig. 6. This check is important to prevent "over-fitting" the NN.

The second test is performed after training. We check the overall consistency of response of the NNs and our modeling of signal and background on the preliminary data sample, collected during Run 1. Figure 7 shows the NN outputs for the sum of signal and background models (central curve with gray error band) and data (triangles). Based on this agreement between the model and data, we conclude that our model describes signal and background, and that the response of the NN is reliable (e.g., there is no over-fitting of the NN).

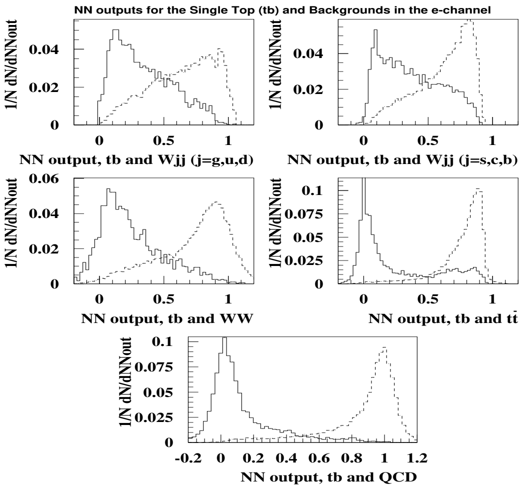

FIGURE 5. NN outputs of trained signal/background pairs. The signal peak lies near one, and the background peak near zero.

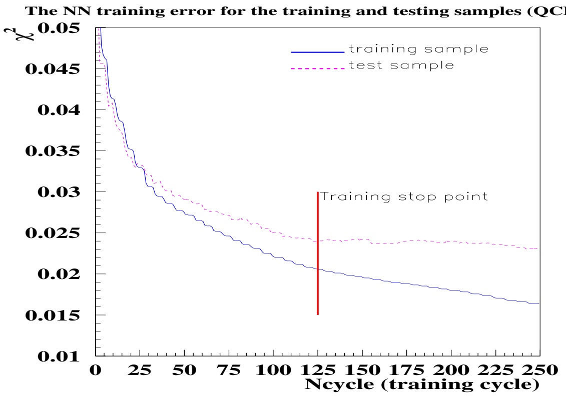

FIGURE 6. Prevention of the over-fitting problem. χ^2 is the error function.

CONCLUSION

After performing these checks, the NNs are ready for the final implementation of cutoffs, after which we can count the expected and observed numbers of events. For our chosen set of cutoffs on the 5 NN outputs, we obtain the following final overall NN efficiencies:

Signal Eff. = Eff.Preprocessing.Cuts $\times$ Eff.NN.Cuts =
$$78\% \times 21\% = 16\%$$
Background Eff. = Eff.Preprocessing.Cuts $\times$ Eff.NN.Cuts =
$$27\% \times 2.7\% = 0.73\%$$

We have applied this analysis technique to a search of DØ data for single top quark production, and the results are being prepared for publication.

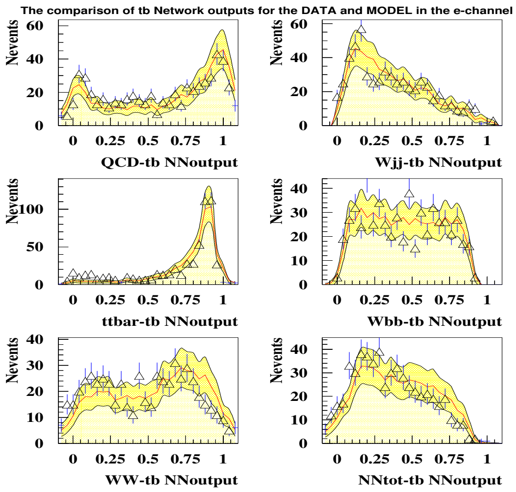

FIGURE 7. Comparison of the NN outputs for the complete signal+background model and DØ Run 1 data.

REFERENCES

1. F. Abe *et al.* (CDF Collaboration), Phys. Rev. Lett. **74**, 2626 (1995); S. Abachi *et al.* (DØ Collaboration), Phys. Rev. Lett. **74**, 2632 (1995).
2. B. Abbot *et al.* (DØ Collaboration), to appear in Phys. Rev. D, hep-ex/0008024; P. Savard, to appear in the proceedings of the Meeting of the Division of Particles and Fields of the American Physical Society, Columbus, OH, August 2000
3. M.C. Smith and S. Willenbrock, Phys. Rev. D **54**, 6696 (1996)
4. T. Stelzer, Z. Sullivan, and S. Willenbrock, Phys. Rev. D **56**, 5919 (1997)
5. E. Boos, L. Dudko and T. Ohl, Eur. Phys. J. C **11**, 473-484 (1999); L. Dudko in the Proceedings of AIHENP 99; E. Boos, L. Dudko, DØNote 3612
6. MLPfit v1.40, J. Schwindling, B. Mansoulie, `http://home.cern.ch/~schwind/MLPfit.html`